

Testing for the significance of functional covariates in regression models

Samuel Maistre* and Valentin Patilea†

Abstract

Regression models with a response variable taking values in a Hilbert space and hybrid covariates are considered. This means two sets of regressors are allowed, one of finite dimension and a second one functional with values in a Hilbert space. The problem we address is the test of the effect of the functional covariates. This problem occurs for instance when checking the goodness-of-fit of some regression models for functional data. The significance test for functional regressors in nonparametric regression with hybrid covariates and scalar or functional responses is another example where the core problem is the test on the effect of functional covariates. We propose a new test based on kernel smoothing. The test statistic is asymptotically standard normal under the null hypothesis provided the smoothing parameter tends to zero at a suitable rate. The one-sided test is consistent against any fixed alternative and detects local alternatives à la Pitman approaching the null hypothesis. In particular we show that neither the dimension of the outcome nor the dimension of the functional covariates influences the theoretical power of the test against such local alternatives. Simulation experiments and a real data application illustrate the performance of the new test with finite samples.

Keywords. Regression, goodness-of-fit test, functional data, U -statistics

*Corresponding author. CREST (Ensai) & IRMAR (UEB), France; samuel.maistre@ensai.fr

†CREST (Ensai) & IRMAR (UEB), France; patilea@ensai.fr

I Introduction

Let $(\mathcal{H}_1, \langle \cdot, \cdot \rangle_{\mathcal{H}_1})$ and $(\mathcal{H}_2, \langle \cdot, \cdot \rangle_{\mathcal{H}_2})$ denote two possibly different Hilbert spaces. The main examples of Hilbert spaces we have in mind are \mathbb{R}^p , for some $p \geq 1$, and $L^2[0, 1]$, the space of squared integrable real-valued functions defined on the unit interval.

Consider the random variables $U \in \mathcal{H}_1$ and $W \in \mathcal{H}_2$ and let Z be a column random vector in \mathbb{R}^q , $q \geq 0$. By convention, $q = 0$ means that Z is a constant. Let (U_i, Z_i, W_i) , $1 \leq i \leq n$, denote a sample of independent copies of (U, Z, W) . The statistical problem we consider is the test of the hypothesis

$$\mathbb{E}[U \mid Z, W] = 0 \quad \text{a.s.}, \quad (\text{I.1})$$

against a general alternative like $\mathbb{P}(\mathbb{E}[U \mid Z, W] = 0) < 1$. This type of problem occurs in many model check problems.

Consider the random variables $U \in \mathcal{H}_1$, $\widetilde{W} \in \mathcal{H}_2$. For illustration, suppose that U is centered. Consider the problem of *testing the effect of the functional variable \widetilde{W}* , that is testing the condition $\mathbb{E}[U \mid \widetilde{W}] = 0$. Patilea et al. (2012b) proposed a test procedure based on projections into finite dimension subspaces of \mathcal{H}_2 . Their test statistic is somehow related to a Kolmogorov-Smirnov statistic in a finite dimension space with the dimension growing with the sample size. Here we propose an alternative route that avoids optimization in high dimension. Let $Z = \langle \widetilde{W}, \phi_1 \rangle_{\mathcal{H}_2}$ where ϕ_1 is an element of an orthonormal basis of \mathcal{H}_2 . Suppose that Z admits a density with respect to the Lebesgue measure. The basis of \mathcal{H}_2 could be the one given by the functional principal components which in general has to be estimated from the data. In such a case, the sample of Z_i 's has to be estimated too. Let $W = \widetilde{W} - \langle \widetilde{W}, \phi_1 \rangle_{\mathcal{H}_2} \phi_1$. Then, testing $\mathbb{E}[U \mid \widetilde{W}] = 0$ is nothing but testing condition (I.1).

Aneiros-Pérez and Vieu (2006) introduced the *semiparametric functional partially linear models* as an extension of the partially linear model to functional data. Such model writes as

$$Y = Z^\top \beta + m(W) + U, \quad \text{with} \quad \mathbb{E}[U \mid Z, W] = 0 \quad \text{a.s.},$$

where Y is a scalar response and Z is a q -dimension vector of random covariates, W is a random variable taking values in a functional space, typically $L^2[0, 1]$. The column vector of q coefficients β and the function $m(\cdot)$ have to be estimated. Before estimating $m(\cdot)$ nonparametrically, one should first check the significance of the variable W which means exactly testing condition (I.1). In this example, the variable U is not observed and the sample U_1, \dots, U_n could be estimated by the residuals of the linear fit of Y given Z . The estimation error for the sample of U is of rate $O_{\mathbb{P}}(n^{-1/2})$ and could be easily proved to be negligible for our test.

Other examples of regression model checks that lead to a problem like (I.1) are the functional linear regression with scalar or functional responses, quadratic functional regression, generalized functional regression, *etc.* See for instance Horváth and Kokoszka (2012) for a recent panorama on the functional regression models. In such situations one has to estimate the sample U_1, \dots, U_n from the functional regression model considered. The estimation error is in general larger than the parametric rate $O_{\mathbb{P}}(n^{-1/2})$, but one can

still show that, under reasonable conditions, it remains negligible for the test purposes. See Patilea et al. (2012b) for a related framework.

Another example, related to the problem of testing the effect of a functional variable, is the *variable selection in functional nonparametric regression with functional responses*. Regression models for functional responses are now widely used, see for instance Faraway (1997). Two situations were studied: finite and infinite dimension covariates; see Ramsay and Silverman (2005), Ferraty et al. (2011), Ferraty et al. (2012). Consider the hybrid case with both finite and infinite dimension covariates. An important question is the significance of the functional covariates. In a more formal way, let $Y \in \mathcal{H}_1$ be the response and let $Z \in \mathbb{R}^q$ and $W \in \mathcal{H}_2$ denote the covariates. Then the problem is to test the equality

$$\mathbb{E}[Y \mid Z, W] = \mathbb{E}[Y \mid Z].$$

Let $U = Y - \mathbb{E}[Y \mid Z]$. Then the problem becomes to test whether $\mathbb{E}[U \mid Z, W] = 0$ almost surely, that is the condition (I.1). Again the sample of the variable U is not observed and has to be estimated by the residuals of the nonparametric regression of Y given Z . See also Lavergne et al. (2014) for a related procedure.

As a last example where a condition like (I.1) occurs consider the problem of testing the independence between a random variable Y and a functional spaced valued variable \widetilde{W} . Without loss of generality, one could suppose that Y takes values in the unit interval. Define $U(t) = \mathbf{1}\{Y \leq t\} - \mathbb{P}(Y \leq t)$, $t \in [0, 1]$, that is centered and belongs to $L^2[0, 1]$. The independence between Y and \widetilde{W} is equivalent to the condition $\mathbb{E}[U \mid \widetilde{W}] = 0$. Conditional independence of Y and a functional random variable given some finite random vector Z could be also tested. It suffices to define $U(t)$ by centering with the conditional probability of the event $\{Y \leq t\}$ given Z and to check a condition like (I.1).

To our best knowledge the statistical problem we address in this work was very little investigated in full generality. Chiou and Müller (2007) and Kokoszka et al. (2008) investigated the problem of goodness-of-fit with functional responses. Chiou and Müller (2007) considered plots of functional principal components (FPC) scores of the response and the covariate. They also used residuals versus fitted values FPC scores plots. However, such two dimension plots could not capture all types of effects of the covariate on the response. Kokoszka et al. (2008) used the response and covariate FPC scores to build a test statistic with χ^2 distribution under the null hypothesis of no linear effect. See also the textbook Horváth and Kokoszka (2012). Again, by construction, such tests cannot detect any nonlinear alternative. The goodness-of-fit or no-effect against nonparametric alternatives has been recently explored in functional data context. In the case of scalar response, Delsol et al. (2011) proposed a testing procedure adapted from the approach of Härdle and Mammen (1993). Their procedure involves smoothing in the functional space and requires quite restrictive conditions. Patilea et al. (2012a) and García-Portugués et al. (2012) proposed alternative nonparametric goodness-of-fit tests for scalar response and functional covariate using projections of the covariate. Patilea et al. (2012b) extended the idea to functional responses and seems to be the only contribution allowing for functional responses. Such projection-based methods are less restrictive and perform well in applications. However, they require a search for the most suitable projec-

tion and this may involve optimization in high dimension.

The paper is organized as follows. In section II we introduce our testing approach, while in section III we provide the asymptotic analysis. The asymptotically standard normal critical values and the consistency of the test are derived. The application to goodness-of-fit tests of functional data models is discussed. The extension to the case of estimated covariates is presented in section IV. This allows in particular for an estimated basis in the infinite-dimensional space of the functional covariate. Section V presents some empirical evidence on the performances of our test and comparisons with existing procedures. The proofs and some technical lemmas are relegated to the Appendix.

II The method

Let us first introduce some notation. Let $\{\phi_1, \phi_2, \dots\}$ be some orthonormal basis of \mathcal{H}_2 that for the moment is supposed to be fixed. In section IV we consider the case of a data-driven basis. For simplicity and without any loss of generality in the following, assume hereafter that $\mathbb{E}(W) = 0$. Then we can decompose $W = \sum_{k \geq 1} \langle W, \phi_k \rangle_{\mathcal{H}_2} \phi_k$ and the norm of W satisfies the relationship $\|W\|_{\mathcal{H}_2}^2 = \sum_{k \geq 1} \langle W, \phi_k \rangle_{\mathcal{H}_2}^2$. Let us note that

$$\mathbb{E}(U \mid Z, W) = \mathbb{E}(U \mid Z, \langle W, \phi_1 \rangle_{\mathcal{H}_2}, \langle W, \phi_2 \rangle_{\mathcal{H}_2}, \dots).$$

Next, for any positive integer p , let

$$W_{\underline{p}} = (\langle W, \phi_1 \rangle_{\mathcal{H}_2}, \dots, \langle W, \phi_p \rangle_{\mathcal{H}_2})^\top.$$

For a function l , let $\mathcal{F}[l]$ denote the Fourier Transform of l . Let K be a multivariate kernel defined on \mathbb{R}^q such that $\mathcal{F}[K] > 0$ and $\varphi(s) = \exp(-\|s\|^2/2)$, $\forall s \in \mathbb{R}^p$, where here $\|\cdot\|$ is the Euclidean norm in \mathbb{R}^p . Many univariate kernels satisfy the positive Fourier Transform condition, for instance the gaussian, triangle, Student and logistic densities. To obtain a multivariate kernel with positive Fourier Transform it suffice to consider a multiplicative kernel with positive Fourier Transform univariate kernels.

II.1 The idea behind the testing method

The new procedure proposed below is motivated by the following facts. First, if (U_1, Z_1, W_1) and (U_2, Z_2, W_2) are independent copies of (U, Z, W) , for any positive function $\omega(\cdot)$ and any $h > 0$ and p positive integer, by the Inverse Fourier Transform formula,

$$\begin{aligned} I_p(h) &= \mathbb{E} \left[\langle U_1, U_2 \rangle_{\mathcal{H}_1} \omega(Z_1) \omega(Z_2) h^{-q} K((Z_1 - Z_2)/h) \varphi(W_{1,\underline{p}} - W_{2,\underline{p}}) \right] \\ &= \mathbb{E} \left[\langle U_1, U_2 \rangle_{\mathcal{H}_1} \omega(Z_1) \omega(Z_2) \int_{\mathbb{R}^q} e^{2\pi i t^\top (Z_1 - Z_2)} \mathcal{F}[K](th) dt \right. \\ &\quad \left. \times \int_{\mathbb{R}^p} e^{2\pi i s^\top (W_{1,\underline{p}} - W_{2,\underline{p}})} \mathcal{F}[\varphi](s) ds \right] \\ &= \int_{\mathbb{R}^q} \int_{\mathbb{R}^p} \left\| \mathbb{E} \left[\mathbb{E}[U \mid Z, W_{\underline{p}}] \omega(Z) e^{-i\{t^\top Z + s^\top W_{\underline{p}}\}} \right] \right\|_{\mathcal{H}_1}^2 \mathcal{F}[K](th) \mathcal{F}[\varphi](s) dt ds. \end{aligned}$$

By the properties of the Fourier Transform and the conditions $\mathcal{F}[\varphi], \mathcal{F}[K] > 0$ (and $\omega > 0$), for any $h > 0$ and p the real number $I_p(h)$ is nonnegative and

$$\mathbb{E}(U \mid Z, W_p) = 0 \text{ a.s.} \Leftrightarrow I_p(h) = 0.$$

Second, by a martingale convergence argument with respect to p , it follows that

$$\mathbb{E}(U \mid Z, W) = 0 \text{ a.s.} \Leftrightarrow \mathbb{E}(U \mid Z, W_p) = 0 \text{ a.s.} \quad \forall p \in \{1, 2, \dots\}.$$

These intuitions are formalized in the following fundamental lemma, up to some technical modification. In the following a_1, a_2, \dots is a fixed sequence of positive real numbers. For any sequences $u = \{u_k\}_{k \geq 1}$, $v = \{v_k\}_{k \geq 1}$, let

$$\langle u, v \rangle_{\mathcal{A}} = \sum_{k \geq 1} a_k u_k v_k \quad \text{and} \quad \|u\|_{\mathcal{A}}^2 = \sum_{k \geq 1} a_k u_k^2, \quad (\text{II.1})$$

whenever the series converge.

Lemma 2.1 *Assume that $\omega(\cdot) > 0$, $\mathcal{F}[K] > 0$, $\mathcal{F}[K]$ is integrable and $\mathbb{E}(\omega^2(Z) \|U\|_{\mathcal{H}_1}^2) < \infty$, $\mathbb{E}(\|W\|_{\mathcal{H}_2}^2) < \infty$. Assume that $\sum_{k \geq 1} a_k < \infty$ and let*

$$I(h) = \mathbb{E} \left[\langle U_1, U_2 \rangle_{\mathcal{H}_1} \omega(Z_1) \omega(Z_2) h^{-q} K((Z_1 - Z_2)/h) \exp(-\|W_1 - W_2\|_{\mathcal{A}}^2/2) \right], \quad h > 0.$$

Then, for any $h > 0$ we have

$$\mathbb{E}(U \mid Z, W) = 0 \text{ a.s.} \Leftrightarrow I(h) = 0.$$

The reason for introducing a sequence $\{a_k\}$ with convergent partial sums is technical. It allows for an inverse Fourier Transform formula in infinite-dimensional Hilbert spaces. In the remark following Theorem 3.1 we argue that considering the weighted norm $\|\cdot\|_{\mathcal{A}}$ is not restrictive.

The idea behind the new approach we propose is to build a test statistic using an approximation of $I(h)$. Moreover, we will let h tend to zero in order to obtain an asymptotically pivotal test statistic with standard gaussian critical values. A convenient choice of the function $\omega(\cdot)$ will allow to simplify this task. As explained below, in many examples one could simply take $\omega(\cdot) \equiv 1$.

II.2 The test statistics

To estimate $I(h)$ using the i.i.d. sample (U_i, Z_i, W_i) , $1 \leq i \leq n$, we consider the U -statistic

$$I_n(h) = \frac{1}{n(n-1)h^q} \sum_{1 \leq i \neq j \leq n} \langle U_i \omega(Z_i), U_j \omega(Z_j) \rangle_{\mathcal{H}_1} K_{ij}(h) \varphi_{ij},$$

where

$$K_{ij}(h) = K((Z_i - Z_j)/h), \quad \varphi_{ij} = \exp(-\|W_i - W_j\|_{\mathcal{A}}^2/2). \quad (\text{II.2})$$

The variance of $I_n(h)$ could be estimated by

$$v_n^2(h) = \frac{2}{n^2(n-1)^2h^{2q}} \sum_{1 \leq i \neq j \leq n} \langle U_i\omega(Z_i), U_j\omega(Z_j) \rangle_{\mathcal{H}_1}^2 K_{ij}^2(h) \varphi_{ij}^2.$$

Then, the test statistic is

$$T_n = \frac{I_n(h)}{v_n(h)}. \quad (\text{II.3})$$

When the $U_i\omega(Z_i)$'s need to be estimated, the test statistics becomes

$$\hat{T}_n = \frac{\hat{I}_n(h)}{\hat{v}_n(h)}, \quad (\text{II.4})$$

where

$$\begin{aligned} \hat{I}_n(h) &= \frac{1}{n(n-1)h^q} \sum_{1 \leq i \neq j \leq n} \left\langle \widehat{U_i\omega(Z_i)}, \widehat{U_j\omega(Z_j)} \right\rangle_{\mathcal{H}_1} K_{ij}(h) \varphi_{ij}, \\ \hat{v}_n^2(h) &= \frac{2}{n^2(n-1)^2h^{2q}} \sum_{1 \leq i \neq j \leq n} \left\langle \widehat{U_i\omega(Z_i)}, \widehat{U_j\omega(Z_j)} \right\rangle_{\mathcal{H}_1}^2 K_{ij}^2(h) \varphi_{ij}^2. \end{aligned}$$

and the $\widehat{U_i\omega(Z_i)}$ are some estimates of the $U_i\omega(Z_i)$'s.

In the example on testing the effect of a functional variable the U_i are supposed observed so that T_n could be used. For the semiparametric functional partially linear models, to build \hat{T}_n it is convenient to take the $\omega(Z_i)$ constant equal to 1 while the \hat{U}_i will be the residuals of the linear model with response Y and covariate vector $Z \in \mathbb{R}^q$. In the other examples of functional regression models mentioned above (functional linear regression with scalar or functional responses, quadratic functional regression, generalized functional regression, *etc.*), it is convenient to set all $\omega(Z_i)$ equal to 1 and take the \hat{U}_i 's to be the residuals of the functional regression model. Below we will provide an example of argument for showing that, under suitable assumptions, replacing the U_i 's by the \hat{U}_i 's does not change the asymptotic behavior of our test statistics. Next, for variable selection in functional nonparametric regression with functional responses one can use \hat{T}_n and a convenient choice is $\omega(\cdot)$ equal to the density of Z and

$$\widehat{U_i\omega(Z_i)} = \frac{1}{n-1} \sum_{k \neq i} (Y_i - Y_k) \frac{1}{g^q} L_{ik}(g),$$

where L is another kernel, $L_{ik}(g) = L((Z_i - Z_k)/g)$ and g is a bandwidth converging to zero at a suitable rate. Showing that the estimation error of the $U_i\omega(Z_i)$'s is negligible for the testing purpose requires more complicated technical assumptions but could be obtained along the lines of the results of Lavergne et al. (2014). However, such an investigation is left for future work. Finally, for testing the independence between a $[0, 1]$ -valued random variable Y and a $L^2[0, 1]$ -valued random variable, one could take $\omega(\cdot) \equiv 1$ and define $\hat{U}_i = \mathbf{1}\{Y_i \leq t\} - n^{-1} \sum_{j=1}^n \mathbf{1}\{Y_j \leq t\}$.

III Asymptotic theory

In this section we investigate the asymptotic properties of T_n under the null hypothesis (I.1) and under a sequence of alternative hypothesis. When the $U_i\omega(X_i)$'s have to be estimated, the idea is to show that the difference $\widehat{T}_n - T_n$ is asymptotically negligible under suitable model assumptions. This aspect is investigated in section III.3 below.

III.1 The asymptotic critical values

Under mild technical conditions we show that the test statistic is asymptotically standard normal under the null hypothesis $\mathbb{E}[U \mid Z, W] = 0$ a.s.

Assumption D

- (a) *The random vectors $(U_1, Z_1, W_1), \dots, (U_n, Z_n, W_n)$ are independent draws from the random vector $(U, Z, W) \in \mathcal{H}_1 \times \mathbb{R}^q \times \mathcal{H}_2$ that satisfies $\mathbb{E}\|U\omega(Z)\|_{\mathcal{H}_1}^4 < \infty$.*
- (b) *(i) The vector Z admits a density f_Z that is either bounded or satisfies the condition $\int_{\mathbb{R}^q} |\mathcal{F}[f_Z]|^{2-a}(t)dt < \infty$ for some $a \in (0, 1]$.*
(ii) The functional covariate satisfies $\mathbb{E}[\|W\|_{\mathcal{H}_2}^2] < \infty$.
(iii) The norm $\|\cdot\|_{\mathcal{A}}$ is defined like in equation (II.1) with a positive sequence $\{a_k\}$ such that $\sum_{k \geq 1} a_k < \infty$.
- (c) $\exists \underline{\sigma}^2, C > 0$ and $\nu > 2$ such that:
 - (i) $0 < \underline{\sigma}^2 \leq \mathbb{E}[\langle U_1\omega(Z_1), U_2\omega(Z_2) \rangle_{\mathcal{H}_1}^2 \mid (Z_1, W_1), (Z_2, W_2)]$ almost surely;
 - (ii) $\mathbb{E}[\|U\omega(Z)\|_{\mathcal{H}_1}^\nu \mid Z, W] \leq C$ almost surely.

Assumption K

- (a) *The kernel K is multiplicative kernel in \mathbb{R}^q , that is $K(z_1, \dots, z_q) = \tilde{K}(z_1) \cdots \tilde{K}(z_q)$ where \tilde{K} is a symmetric density of bounded variation on real line. Moreover the Fourier Transform $\mathcal{F}[\tilde{K}]$ is positive and integrable.*
- (b) $h \rightarrow 0$ and $nh^q / \ln n \rightarrow \infty$.

Theorem 3.1 *Under the Assumptions D and K the test statistic T_n converges in law to a standard normal, provided the hypothesis (I.1) is true. Consequently, the test given by $\mathbb{I}(T_n \geq z_{1-\alpha})$, with z_α the $(1 - \alpha)$ -quantile of the standard normal distribution, has asymptotic level α .*

Remark 1. Let us comment on Assumption-(D)-(ii,iii). Suppose that the functional covariate satisfies $\mathbb{E}[\langle W, \phi_k \rangle_{\mathcal{H}_2}^2] \sim k^{-\beta}$, $\forall k \geq 1$, for some $\beta > 0$. If $\beta > 2$ one could use directly $\|\cdot\|_{\mathcal{H}_2}$ instead of $\|\cdot\|_{\mathcal{A}}$ to build the test. Indeed, taking $a_k \sim k^{-\beta/2}$ and replacing

W by $\check{W} = \sum_{k \geq 1} b_k^{-1} \langle W, \phi_k \rangle_{\mathcal{H}_2} \phi_k$, with $b_k = a_k^{1/2}$, one would have $\mathbb{E}(U \mid Z, W) = \mathbb{E}(U \mid Z, \check{W})$, $\check{W} \in \mathcal{H}_2$ and $\|\check{W}_i - \check{W}_j\|_{\mathcal{A}} = \|W_i - W_j\|_{\mathcal{H}_2}$.

In the case $\beta = 2$, one could take $a_k \sim k^{-1}(\ln k)^{-(1+\varepsilon)}$, for some $\varepsilon > 0$, and replace W by $\check{W} = \sum_{k \geq 1} a_k^{-1/2} (\ln k)^{-(1+\varepsilon)} \langle W, \phi_k \rangle_{\mathcal{H}_2} \phi_k$. In this case one still has $\mathbb{E}(U \mid Z, W) = \mathbb{E}(U \mid Z, \check{W})$ and $\check{W} \in \mathcal{H}_2$. However, $\|\check{W}_i - \check{W}_j\|_{\mathcal{A}}$ and $\|W_i - W_j\|_{\mathcal{H}_2}$ are no longer equal but in general remain close. Our simulation experiments reveal that in many situations where $\beta = 2$ one could confidently use $\|W_i - W_j\|_{\mathcal{H}_2}$ instead of $\|\check{W}_i - \check{W}_j\|_{\mathcal{A}}$ to build the test.

Finally, with suitable choices, our setup covers also the range $0 < \beta < 2$. When $1 < \beta < 2$, one can set $a_k \sim k^{-1} \ln^{-(1+\varepsilon)} k$ and work with $\|W_i - W_j\|_{\mathcal{A}}$. For the case $0 < \beta \leq 1$ one could transform W in $\check{W} = \sum_{k \geq 1} b_k^{-1} \langle W, \phi_k \rangle_{\mathcal{H}_2} \phi_k$ with $b_k = k^{(1-\beta)/2} \ln^{(1+\varepsilon)/2} k$, and take $a_k \sim k^{\beta-2} \ln^{-2(1+\varepsilon)} k$. The test is then built with $\|\check{W}_i - \check{W}_j\|_{\mathcal{A}}$.

In summary, Assumption-(D)-(ii,iii) represent mild conditions that are satisfied directly, or after simple modifications of the covariate W , in most situations. ■

III.2 The consistency of the test

Let (U_i^0, Z_i, W_i) , $i \geq 1$, i.i.d. such that $\mathbb{E}(U_i^0 \mid Z_i, W_i) = 0$ almost surely. Here we show that our test is consistent against any fixed alternative and detect Pitman alternatives

$$H_{1n} : U_{in} = U_i^0 + r_n \delta(Z_i, W_i), \quad 1 \leq i \leq n, \quad n \geq 1,$$

with probability tending to 1, provided that the rate of decrease of the sequence $\{r_n\}$ satisfies some conditions. These conditions are the same as for nonparametric checks of parametric regression models with finite dimension covariates.

Theorem 3.2 *Under the Assumptions D holds for the (U_i^0, Z_i, W_i) 's and Assumption K holds too. Suppose that $\delta(\cdot, \cdot)$ and r_n defining the sequence of H_{1n} satisfy the conditions $\mathbb{E}[\delta(Z, W)] = 0$, $0 < \mathbb{E}[\|\delta(Z, W)\omega(Z)\|_{\mathcal{H}_1}^4] < \infty$ and $r_n^2 n h^{q/2} \rightarrow \infty$. Then the test based on T_n is consistent against the sequence of alternatives H_{1n} .*

The zero mean condition for $\delta(\cdot)$ keeps U_{in} of zero mean under the alternative hypotheses H_{1n} . The proof is based on standard arguments and is relegated to the appendix.

III.3 Goodness-of-fit test

In this section we provide some guidelines on how our test could be used for testing the goodness-of-fit of functional data models. The detailed investigation of specific situations depend on the model and could not be considered in a unified framework.

In many situations, the U_i 's are not observed and one has to replace them by some \hat{U}_i obtained as residuals of some models. In this case one cannot build T_n and has to work with the statistic \hat{T}_n defined in equation (II.4) instead. In section V we use some simulation experiments to show that our test could still perform well in such situations,

especially with a bootstrap correction, as described in the following, when the sample size is not large enough.

From the theoretical point of view, one shall expect that the asymptotic standard normal critical values are still valid and the test is still consistent, provided that the difference $\widehat{U}_i - U_i$ could be controlled in a suitable way. Indeed, using the notation from section (II.2) and considering the simple case where $\omega(\cdot) \equiv 1$, we can write

$$\begin{aligned}\widehat{I}_n(h) &= I_n(h) \\ &\quad + \frac{2}{n(n-1)h^q} \sum_{1 \leq i \neq j \leq n} \left\langle U_i, \widehat{U}_j - U_j \right\rangle_{\mathcal{H}_1} K_{ij}(h) \varphi_{ij} \\ &\quad + \frac{1}{n(n-1)h^q} \sum_{1 \leq i \neq j \leq n} \left\langle \widehat{U}_i - U_i, \widehat{U}_j - U_j \right\rangle_{\mathcal{H}_1} K_{ij}(h) \varphi_{ij} \\ &\stackrel{\text{def}}{=} I_n(h) + R_{1n}(h) + R_{2n}(h).\end{aligned}$$

Next, one has to control $\widehat{U}_i - U_i$ and hence $R_{1n}(h)$ and $R_{2n}(h)$ and this strongly depends on the specific model considered. Many functional data models would allow to show that $R_{1n}(h)$ and $R_{2n}(h)$ are negligible under reasonable conditions in the model (regularity conditions on the model parameter and the functional covariate W) and for suitable rates of the bandwidth. For instance, Patilea et al. (2012a) investigated in detail the case of linear model with scalar responses. Their investigation could be adapted to our test and obtain similar conclusions. In the case of a functional linear model with $L^2[0, 1]$ responses and finite and infinite dimension covariates one would observe a sample of (Y, Z, W) where

$$Y(t) = Z^\top \beta + \langle \xi(t, \cdot), W \rangle_{\mathcal{H}_2} + U(t), \quad t \in [0, 1].$$

Since β is expected to be estimated at parametric rate, the control of $\widehat{U}_i - U_i$ would depend on the conditions on the rate of convergence of $\widehat{\xi}(\cdot, \cdot)$, the estimate of $\xi(\cdot, \cdot)$. Under suitable but mild conditions, one could expect the rate of $R_{1n}(h)$ to be of order $n^{-1}h^{-q/2}$ times the norm of $\widehat{\xi}(\cdot, \cdot) - \xi(\cdot, \cdot)$, while the rate of $R_{2n}(h)$ to given by the square of the norm of $\widehat{\xi}(\cdot, \cdot) - \xi(\cdot, \cdot)$. Meanwhile, the rate of $I_n(h)$ is $O_{\mathbb{P}}(n^{-1}h^{-p/2})$. The required restrictions on the bandwidth to preserve the asymptotic standard normal critical values follow. Let us point out that slower rates for the norm of $\widehat{\xi}(\cdot, \cdot) - \xi(\cdot, \cdot)$ will require faster decreases for h , and this will result in a loss of power against sequences of local alternatives.

III.4 Bootstrap critical values

To correct the finite sample critical values let us propose a simple wild bootstrap procedure. The bootstrap sample, denoted by U_i^* , $1 \leq i \leq n$, is defined as $U_i^* = \zeta_i U_i$, $1 \leq i \leq n$, where ζ_i , $1 \leq i \leq n$ are independent random variables following the two-points distribution proposed by Mammen (1993). That means $\zeta_i = -(\sqrt{5} - 1)/2$ with probability $(\sqrt{5} + 1)/(2\sqrt{5})$ and $\zeta_i = (\sqrt{5} + 1)/2$ with probability $(\sqrt{5} - 1)/(2\sqrt{5})$. A bootstrap test statistic T_n^* is built from a bootstrap sample as was the original test statistic. When this scheme is repeated many times, the bootstrap critical value $z_{1-\alpha, n}^*$ at level α is the

empirical $(1 - \alpha)$ -th quantile of the bootstrapped test statistics. The asymptotic validity of this bootstrap procedure is guaranteed by the following result. It states that the bootstrap critical values are asymptotically standard normal under the null hypothesis and under the alternatives like in section III.2. The proof could be obtained by rather standard modifications of the proof of Theorem 3.1 and hence will be omitted.

Theorem 3.3 *Suppose that the conditions of Theorem 3.2 hold true, in particular in the case $r_n \equiv 0$. Then*

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(T_n^* \leq x \mid U_1, Z_1, W_1, \dots, U_n, Z_n, W_n) - \mathbb{P}(T_n \leq x)| \rightarrow 0, \quad \text{in probability.}$$

IV The error in covariates case

In this section we show that our testing procedure extends to the case where the covariates are observed with error. In some applications, the observations Z_i and W_i are not directly observed but could be estimated by some \widehat{Z}_i and \widehat{W}_i computed from the data. To better illustrate the methodology, let us focus on the test for the effect of a functional variable. For this reason, in this section let us take $q = 1$, $Z = \langle \widetilde{W}, \phi_1 \rangle_{\mathcal{H}_2}$ and $W = \widetilde{W} - \langle \widetilde{W}, \phi_1 \rangle_{\mathcal{H}_2} \phi_1$, where $\widetilde{W} \in \mathcal{H}_2$ and ϕ_1, ϕ_2, \dots are the elements of an orthonormal basis in \mathcal{H}_2 .

In functional data analysis where usually $\mathcal{H}_2 = L^2[0, 1]$ the choice of the basis is a key point. The statistician would likely prefer a basis allowing an accurate representation of \widetilde{W} with a minimal number of basis elements. A commonly used basis is given by the eigenfunctions of the covariance operator \mathcal{K} that is defined by $(\mathcal{K}v)(\cdot) = \int \mathcal{K}(\cdot, s)v(s)ds$, $v \in L^2[0, 1]$, where \widetilde{W} is supposed to satisfy $\int \mathbb{E}(\widetilde{W}^2(t))dt < \infty$ and

$$\mathcal{K}(t, s) = \mathbb{E}[\{\widetilde{W}(t) - \mathbb{E}(\widetilde{W}(t))\}\{\widetilde{W}(s) - \mathbb{E}(\widetilde{W}(s))\}]$$

is supposed positive definite. Let $\theta_1, \theta_2, \dots$ denote the eigenvalues of \mathcal{K} and let $\mathcal{R} = \{\phi_1, \phi_2, \dots\}$ be the corresponding basis of eigenfunctions that are usually called the functional principal components (FPC). The FPC orthonormal basis provide optimal (with respect to the mean-squared error) low-dimension representations of \widetilde{W} . See, for instance, Ramsay and Silverman (2005). In most of the applications the FPC are unknown and has to be estimated from

$$(\widehat{\mathcal{K}}v)(t) = \int_{[0,1]} \widehat{\mathcal{K}}(t, s)v(s)ds, \quad t \in [0, 1],$$

where

$$\widehat{\mathcal{K}}(t, s) = n^{-1} \sum_{i=1}^n \left\{ \widetilde{W}_i(t) - n^{-1} \sum_{j=1}^n \widetilde{W}_j(t) \right\} \left\{ \widetilde{W}_i(s) - n^{-1} \sum_{j=1}^n \widetilde{W}_j(s) \right\}. \quad (\text{IV.1})$$

Let $\widehat{\theta}_1, \widehat{\theta}_2, \dots \geq 0$ denote the eigenvalues of $\widehat{\mathcal{K}}$ and let $\widehat{\phi}_1, \widehat{\phi}_2, \dots$ be the corresponding basis of eigenfunctions, that is the estimated FPC. For identification purposes, we adopt

the usual condition $\langle \phi_j, \hat{\phi}_j \rangle \geq 0, \forall j$. Now, we can define $\hat{Z}_i = \langle \widehat{W}_i, \hat{\phi}_1 \rangle_{\mathcal{H}_2}$ and $\widehat{W}_i = \widehat{W}_i - \langle \widehat{W}_i, \hat{\phi}_1 \rangle_{\mathcal{H}_2} \hat{\phi}_1$, the estimates of Z_i and W_i .

Having in mind such types of situations, herein we will suppose that

$$\|\hat{Z}_i - Z_i\| + \|\widehat{W}_i - W_i\|_{\mathcal{H}_2} \leq \Gamma_i \Delta_n, \quad 1 \leq i \leq n, \quad (\text{IV.2})$$

where Γ_i are independent copies of some random variable Γ that depend on X , and Δ_n depend on the data but could be taken the same for all i . For Δ_n and Γ we will suppose

$$\Delta_n = O_{\mathbb{P}}(n^{-1/2}) \quad \text{and} \quad \exists a > 0 \quad \text{such that} \quad \mathbb{E} \exp(a\Gamma) < \infty. \quad (\text{IV.3})$$

Clearly, alternative conditions on the rate of Δ_n and the moments of Γ could be considered, resulting in alternative conditions on the bandwidths in the statements below. As it will be explained below, the conditions (IV.3) are convenient for the example of \hat{Z}_i and \widehat{W}_i obtained from estimated FPC basis.

Let us introduce some notation

$$\hat{K}_{ij}(h) = K((\hat{Z}_i - \hat{Z}_j)/h), \quad \hat{\varphi}_{ij} = \exp(-\|\widehat{W}_i - \widehat{W}_j\|_{\mathcal{A}}^2/2). \quad (\text{IV.4})$$

Let

$$\tilde{T}_n = \frac{\tilde{I}_n(h)}{\tilde{v}_n(h)}.$$

where $\tilde{I}_n(h)$ and $\tilde{v}_n(h)$ are defined as $I_n(h)$ and $v_n(h)$ in section II.2 but with \hat{Z}_i and \widehat{W}_i instead of Z_i and W_i .

Theorem 4.1 *Suppose that $q = 1$, the Assumptions D-(a), D-(b)-(ii, iii), K-(a) are met and conditions (IV.2) and (IV.3) hold true. Assume one of the following conditions is met:*

1. $nh^4/\ln^2 n \rightarrow \infty$ and $\int_{\mathbb{R}^q} |\mathcal{F}[f_Z]|^{2-a}(t) dt < \infty$ for some $a \in (0, 1]$;
2. $nh^2/\ln^2 n \rightarrow \infty$ and f_Z is bounded;
3. $nh/\ln n \rightarrow \infty$ and f_Z is uniformly continuous.

Then the Theorems 3.1, 3.2 and 3.3 remain valid with the test statistic T_n replaced by \tilde{T}_n .

The proof of Theorem 4.1 is a direct consequence of Lemma 6.1 in the Appendix and is hence omitted.

Let us revisit the problem of the test for the effect of a functional variable, where $Z = \langle \widehat{W}, \phi_1 \rangle_{\mathcal{H}_2}$ and $W = \widehat{W} - \langle \widehat{W}, \phi_1 \rangle_{\mathcal{H}_2} \phi_1$. The conditions on the random variable Z required in Lemma 6.1 are mild conditions satisfied in the common examples of functional covariates considered in the literature. Concerning condition (IV.2), consider the operator norm $\|\mathcal{K}\|_S$ defined by

$$\|\mathcal{K}\|_S^2 = \int \int \sigma^2(t, s) dt ds.$$

Under Assumptions D-(a), D-(b)-(ii), the empirical covariance operator satisfies

$$\|\widehat{\mathcal{K}} - \mathcal{K}\|_S = O_{\mathbb{P}}(1/\sqrt{n}),$$

see for instance Bosq (2000) or Horváth and Kokoszka (2012). On the other hand, suppose that θ_1 , the eigenvalue associated to ϕ_1 , is different from all the others eigenvalues of the operator \mathcal{K} . By Lemma 4.3 in Bosq (2000) or Lemma 2.3 in Horváth and Kokoszka (2012), and the fact that the spectral norm of the operator $\widehat{\mathcal{K}} - \mathcal{K}$ is smaller or equal to $\|\widehat{\mathcal{K}} - \mathcal{K}\|_S$,

$$\|\widehat{\phi}_1 - \phi_1\|^2 \leq \frac{8}{\varsigma^2} \|\widehat{\mathcal{K}} - \mathcal{K}\|_S^2,$$

where ς is the distance between θ_1 and the set $\{\theta_2, \theta_3, \dots\}$ of all the other eigenvalues of \mathcal{K} . Here the eigenvalues of \mathcal{K} are not necessarily ordered, θ_1 could be any eigenvalue separated from all the others. Deduce that condition (IV.2) is guaranteed for instance if there exists $a > 0$ such that

$$\mathbb{E} \exp(a \|\widetilde{W}\|_{\mathcal{H}_2}) < \infty.$$

The exponential moment condition is met if, for instance, \widetilde{W} is a mean-zero Gaussian process defined on the unit interval with $\sup_{t \in [0,1]} \mathbb{E}[\widetilde{W}^2(t)] < \infty$; see chapter A.2 in van der Vaart and Wellner (1996). Moreover, in general, a moment restriction on \widetilde{W} is not restrictive for significance testing. Indeed, if \widetilde{W} does not satisfy such a condition, it suffices to transform \widetilde{W} into some variable $\in \mathcal{H}_2$ such that \widetilde{V} generates the same σ -field and \widetilde{V} satisfies the required moment condition.

V Empirical evidence

In this section we illustrate the empirical performances of our testing procedure. For that purpose, we consider both scalar and functional responses Y . We used an Epanechnikov kernel in our applications, that is $K(x) = 0.75(1 - x^2) \mathbf{1}\{|x| < 1\}$. We calculated φ_{ij} and $\widehat{\varphi}_{ij}$ in two ways: with the norm in the Hilbert space $L^2[0, 1]$ of the covariate and with the norm $\|\cdot\|_{\mathcal{H}}$ proposed in Remark 1 for the case $\beta = 2$.

Below $\langle \cdot, \cdot \rangle$ is the usual inner product on $L^2[0, 1]$, that is $\langle f, g \rangle = \int_0^1 f(t) g(t) dt$. Let $\widehat{\mathcal{K}}$ be the empirical covariance operator defined in (IV.1) and let $\widehat{\theta}_1 \geq \widehat{\theta}_2 \geq \dots \geq 0$ be its eigenvalues and $\widehat{\phi}_1, \widehat{\phi}_2, \dots$ the corresponding eigenfunctions.

V.1 The scalar response case

We simulate data samples of size $n = 40$ using the models

$$Y_i = a + \langle X_i, b \rangle + \delta \langle X_i, b \rangle^2 + U_i, \quad (V.1)$$

$$Y_i = (\lambda_k^{-1} \langle X_i, e_k \rangle^2 - 1) + U_i, \quad 1 \leq i \leq n, \quad (V.2)$$

where X_i is a Wiener process, U_i are independent centered normal variables with variance $\sigma^2 = 1/16$,

$$a = 0 \quad \text{and} \quad b(t) = \sin^3(2\pi t^3), \quad t \in [0, 1].$$

Moreover,

$$e_k(t) = \sqrt{2} \sin((k - 1/2)\pi t), \quad t \in [0, 1],$$

and $\lambda_k = (k - 1/2)^{-2}\pi^{-2}$ and k is some fixed positive integer. The null hypothesis corresponds to $\delta = 0$ while nonnegative δ 's yield quadratic alternatives.

We then estimate b using the functional principal component approach, see e.g., Ramsay and Silverman (2005) and Horváth and Kokoszka (2012). The first five principal components of the X_i s are used so that b is estimated by

$$\widehat{b}(t) = \sum_{j=1}^5 \widehat{b}_j \widehat{\phi}_j(t),$$

where $\widehat{b}_j = \widehat{\theta}_j^{-1} \widehat{g}_j$, $\widehat{g}_j = \langle \widehat{g}, \widehat{\phi}_j \rangle$ with

$$\widehat{g}(t) = \frac{1}{n} \sum_{i=1}^n (Y_i - \overline{Y}_n) (X_i(t) - \overline{X}_n(t))$$

and a by $\widehat{a} = \overline{Y}_n - \langle \overline{X}_n, \widehat{b} \rangle$. The test statistics are built with $q = 1$,

$$\widehat{U}_i = Y_i - \widehat{a} - \langle X_i, \widehat{b} \rangle, \quad \widehat{Z}_i = \frac{\widetilde{Z}_i}{\sqrt{n^{-1} \sum_{j=1}^n \widetilde{Z}_j^2}} \quad \text{and} \quad \widehat{W}_i = \frac{\widetilde{W}_i}{\sqrt{n^{-1} \sum_{j=1}^n \|\widetilde{W}_j\|_{\mathcal{H}_2}^2}},$$

where $\widetilde{Z}_i = \langle X_i, \widehat{\phi}_1 \rangle$ and $\widetilde{W}_i = X_i - \langle X_i, \widehat{\phi}_1 \rangle \widehat{\phi}_1$.

First, we investigate the accuracy of the asymptotic critical values and the effectiveness of the bootstrap correction, with 199 bootstrap samples, for level $\alpha = 10\%$. Several bandwidths are considered, that is $h = cn^{-1/5}$ with $c \in \{2^{k/2}, k = -2, -1, 0, 1, 2\}$. The results of 5000 replications are plotted in the left panel of Figure 1. The normal critical values are quite inaccurate, while the bootstrap corrections are very effective, whatever the considered bandwidth is. The differences between the results for the statistics defined with $\|\cdot\|_{\mathcal{H}_2}$ and those for the statistics defined with $\|\cdot\|_{\mathcal{A}}$ are imperceptible.

Next, we compare our test to the one introduced by Patilea et al. (2012a) (hereafter PSSa) based on projections. The test statistic of PSSa is

$$T_n^{PSSa} = \frac{Q_n(\widehat{\gamma}_n; \widehat{a}, \widehat{b})}{\widehat{v}_n(\widehat{\gamma}_n; \widehat{a}, \widehat{b})}$$

where

$$Q_n(\gamma; \widehat{a}, \widehat{b}) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \widehat{U}_i \widehat{U}_j \frac{1}{h} K(h^{-1} \langle X_i - X_j, \gamma \rangle), \quad \gamma \in \mathbb{R}^p,$$

and $\widehat{v}_n^2(\gamma; \widehat{a}, \widehat{b})$ is an estimation of the variance of $nh^{1/2}Q_n(\gamma; \widehat{a}, \widehat{b})$. Here and in the following, the vector $\gamma = (\gamma_1, \dots, \gamma_p)^\top$ is identified with $\sum_{k=1}^p \gamma_k \phi_k \in L^2[0, 1]$. The value of p is chosen by the statistician. The direction $\widehat{\gamma}_n$ is selected as

$$\widehat{\gamma}_n = \arg \max_{\gamma \in B_p} \left[nh^{1/2}Q_n(\gamma, \widehat{a}, \widehat{b}) / \widehat{v}_n(\gamma, \widehat{a}, \widehat{b}) - \alpha_n \mathbf{1}\{\gamma \neq \gamma_0\} \right],$$

where $B_p \subset \mathcal{S}^p = \{\gamma \in \mathbb{R}^p : \|\gamma\| = 1\}$ is a set of positive Lebesgue measure on \mathcal{S}^p and γ_0 is a privileged direction chosen by the statistician and α_n is a penalty term. Here we follow PSSa and we take $p = 3$ and B_3 as a set of 1200 points on \mathcal{S}^3 , $\gamma_0 = (1, 1, 1)/\sqrt{3}$ and $\alpha_n = 3$.

The results are presented on Figure 2 the null hypothesis (5000 replications) and several alternatives (2500 replications) defined by some positive values of δ . The PSSa statistic is computed with wild bootstrap critical values. The rejection rate for the bootstrap version of our test appears to be better than that based on asymptotic critical values for each considered alternative. Moreover, the results obtained with $\|\cdot\|_{\mathcal{H}_2}$ are better than those obtained with $\|\cdot\|_{\mathcal{A}}$. The PSS1 outperforms our test in terms of power for the setups (V.1) and (V.2) with $k = 2$. This could be explained by the nature of the PSS1 statistic which by construction is powerful against such alternatives. When considering the setup (V.2) with $k = 4$ the power is deteriorates drastically for all the tests. The fourth coordinate $\langle X_i, e_4 \rangle$ being independent of the first three involved in the PSS1 statistic, the empirical power of that statistic is practically equal to the level for any sample size. The empirical power of our statistic improves with the sample size and so confirms the asymptotic results. The plateau for the empirical rejection curves for our test could be explained by an inflated variance on the alternatives, but its level increases with the sample size.

V.2 The case of functional response

Three models with functional Y are considered:

$$Y_i(t) = \delta \times \beta(t) X_i(t) + \epsilon_i(t) \quad (\text{V.3})$$

$$Y_i(t) = \delta \times H(B_i(t)) + \epsilon_i(t) \quad (\text{V.4})$$

$$Y_i(t) = \delta \times \lambda_k^{-1/2} \langle B_i, e_k \rangle + \epsilon_i(t) \quad (\text{V.5})$$

$1 \leq i \leq n$, where X_i and ϵ_i are independent Brownian bridges, B_i is a Brownian motion,

$$\beta(t) = \exp\{-4(t - 0.3)^2\}, \quad t \in [0, 1],$$

$e_k(\cdot)$ and λ_k are defined as in the case of scalar response for some fixed $k \geq 1$, and $H(x) = x^2 - 1$, $x \in \mathbb{R}$. We consider $q = 1$ and the \widehat{Z}_i , \widehat{W}_i and $\widehat{\varphi}_{ij}$ are built like in the case of a scalar response.

We compare our test with the one considered by Patilea et al. (2012b) (hereafter PSSb). Their statistic, let us call it T_n^{PSSb} , which is a variant of T_n^{PSSa} above defined with a different Q_n . That is, in the definition of Q_n the product $\widehat{U}_i \widehat{U}_j$ is replaced by the scalar product $\langle \widehat{U}_i, \widehat{U}_j \rangle$ and $K(h^{-1} \langle X_i - X_j, \gamma \rangle)$ by

$$K_h(h^{-1}[F_{\gamma,n}(\langle X_i, \gamma \rangle) - F_{\gamma,n}(\langle X_j, \gamma \rangle)])$$

where $F_{\gamma,n}$ is the empirical c.d.f. of the sample $\langle X_1, \gamma \rangle, \dots, \langle X_n, \gamma \rangle$, $\gamma \in B_p \subset \mathbb{R}^p$. Following PSS2, in this case we take $p = 3$, B_3 as a set of 1200 points on \mathcal{S}^3 , $\gamma_0 =$

$(1, 1, 1)/\sqrt{3}$ and $\alpha_n = 2$. Moreover, since here we test for the effect, \widehat{U}_i are nothing but the observations Y_i .

We also compare our test with the test of Kokoszka et al. (2008) (hereafter KMSZ) based on the eigenvalues $(\widehat{\gamma}_k)_k$ and $(\widehat{\lambda}_k)_k$ and eigenvectors $(\widehat{u}_k)_k$ and $(\widehat{v}_k)_k$, $1 \leq k \leq n$ of the respective empirical operators

$$\Gamma_n x = \frac{1}{n} \sum_{i=1}^n \langle X_i, x \rangle X_i, \quad \Lambda_n x = \frac{1}{n} \sum_{i=1}^n \langle Y_i, x \rangle Y_i,$$

and also

$$\Delta_n x = \frac{1}{n} \sum_{i=1}^n \langle X_i, x \rangle Y_i,$$

the test statistic being

$$T_n(\tilde{p}, \tilde{q}) = n \sum_{k=1}^{\tilde{p}} \sum_{j=1}^{\tilde{q}} \frac{\langle \Delta_n \widehat{v}_k, \widehat{u}_j \rangle^2}{\widehat{\gamma}_k \widehat{\lambda}_j}.$$

This statistic is asymptotically $\chi^2(\tilde{p}\tilde{q})$ distributed when there is no linear effect of X on Y . We test the “no effect” model on the three setups (V.3), (V.4) and (V.5) using $\widehat{U}_i = Y_i - \bar{Y}_n$. For this we consider the cases $\tilde{p} = 1$, $\tilde{q} = 6$ and $\tilde{p} = 2$, $\tilde{q} = 6$.

Again, we investigate the accuracy of the asymptotic critical values and the bootstrap correction, following the same steps as in the case of scalar Y , this time for 1000 replications under the null hypothesis. We present the results in the right panel of Figure 1. The conclusions are similar to those of the scalar case, that is the asymptotic critical values are rather inaccurate with $n = 40$. The bootstrap correction is quite effective, whatever the considered bandwidth is. The empirical power results for positive deviations δ for the three models considered are presented in Figure 3. They are based on a number of 500 replications of the experiment. The results obtained with $\|\cdot\|_{\mathcal{H}_2}$ are again preferable. One can see that KMSZ and PSS2 perform very well for the concurrent alternative. However, for a linear alternative with $k = 4$, the bootstrap version of our test seems to be the best choice. The good performance of the PSS2 with samples of size $n = 40$ could be explained by a correlation between $\langle B_i, \widehat{\phi}_1 \rangle, \dots, \langle B_i, \widehat{\phi}_3 \rangle$ and $\langle B_i, \widehat{\phi}_4 \rangle$ which approximates $\langle B_i, e_4 \rangle$. This correlation vanishes when n increases resulting in a loss of power for PSS2 test. In this experiment we also studied the effect of larger dimension q with $n = 40$ and the concurrent alternative, equation (V.3), and quadratic alternative, equation (V.4). The results presented in Figure 3 reveals a drastic decrease of power. A possible explanation is that when the first components $\langle X_i, \widehat{\phi}_1 \rangle$ carry enough information on the covariate, the price to pay in terms of power for smoothing in higher dimension could be too high, so that it may be preferable to consider $q = 1$.

V.3 Real data application

The approach proposed in this paper is applied to check the goodness-of-fit of several models for the Canadian weather dataset. This dataset is studied in Ramsay and Silverman

(2005) and is included in the R package fda (<http://www.r-project.org>). The data consist of the daily mean temperature and rain registered in 35 weather stations in Canada. A curve is available for each station, describing the rainfall for each day of the year. This is the functional response. The same type of curve with the temperature is used as functional predictor. Several regression models with functional covariate and functional response have been studied in Ramsay and Silverman (2005), and illustrated with the Canadian weather dataset. The purpose here is to assess the validity of each of the following three models

$$Y_{ij}(t) = \mu(t) + \varepsilon_{ij}(t), \quad (\text{V.6})$$

$$Y_{ij}(t) = \mu(t) + \alpha_j(t) + \varepsilon_{ij}(t), \quad (\text{V.7})$$

$$Y_{ij}(t) = \mu(t) + \alpha_j(t) + \int X_{ij}(s) \xi(s, t) ds + \varepsilon_{ij}(t), \quad (\text{V.8})$$

where $\sum_{j=1}^J \alpha_j(\cdot) \equiv 0$ to ensure identification of models (V.7) and (V.8). The stations are classified in four climatic zones (Atlantic, Pacific, Continental, Arctic) and $Y_{ij}(t)$ represents the logarithm of the rainfall at the station i of the climate zone j on day t , $X_{ij}(t)$ is the temperature at the same station on day t of the year. Since each observation Y_{ij} is observed for the same time design, we just use

$$\widehat{U}_{ij}(\cdot) = Y_{ij}(\cdot) - \overline{Y}_n^{(i)}(\cdot) \text{ and } \widehat{U}_{ij}(\cdot) = Y_{ij}(\cdot) - \overline{Y}_{\cdot j}^{(i)}(\cdot)$$

for models (V.6) and (V.7) respectively. Here we use the notation

$$\overline{A}_n^{(i)} = (n-1)^{-1} \left(-A_{ij} + \sum_{j=1}^J \sum_{k=1}^{n_j} A_{kj} \right)$$

and $\overline{A}_{\cdot j}^{(i)} = (n_j-1)^{-1} \sum_{k \in \{1, \dots, n_j\} \setminus \{i\}} A_{kj}$ represent respectively the leave-one out overall mean and the class j mean for the variable A and the observation i . For the model (V.8), let us notice that

$$\overline{Y}_{\cdot j}^{(i)}(t) = \mu(t) + \alpha_j(t) + \int \overline{X}_{\cdot j}^{(i)}(s) \xi(s, t) ds + \overline{\varepsilon}_{\cdot j}^{(i)}(t)$$

and then

$$\tilde{Y}_{ij}(t) = \int \tilde{X}_{ij} \xi(s, t) ds + \tilde{\varepsilon}_{ij}$$

where $\tilde{A}_{ij} = A_{ij} - A_{\cdot j}^{(i)}$. Thus we construct the functional principal components based on $\left\{ \tilde{X}_{ij}(\cdot), j \in \{1, \dots, J\}, i \in \{1, \dots, n_j\} \right\}$ which leads to $\tilde{X}_{ij}(\cdot) \simeq \sum_{\kappa=1}^K \lambda_{\kappa} c_{ij\kappa} v_{\kappa}(\cdot)$ (where $\sum_{j=1}^J \sum_{i=1}^{n_j} c_{ij\kappa}^2 = 1$, $\|v_{\kappa}(\cdot)\|^2 = 1$ and $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$) and

$$\widehat{U}_{ij}(t) = \tilde{Y}_{ij}(t) - \sum_{\kappa=1}^K c_{ij\kappa} \left(-c_{ij\kappa} \tilde{Y}_{ij}(t) + \sum_{l=1}^J \sum_{k=1}^{n_l} c_{kl\kappa} \tilde{Y}_{kl}(t) \right).$$

Model	$q = 1$	$q = 3$
(V.6)	0	0
(V.7)	0	0.009
(V.8), $K = 12$	0.016	0.403
(V.8), $K = 13$	0.023	0.736
(V.8), $K = 14$	0.015	0.723

Table 1: Bootstrap p -values for modeling Canadian Weather data three different ways and for different smoothing dimension $q \in \{1, 2\}$ and 999 bootstrap samples.

All this leave-one-out feature is used to avoid overfitting and for the choice of $K = 13$, we used the one that minimizes $\sum_{j=1}^J \sum_{i=1}^{n_j} \|\widehat{U}_{ij}(\cdot)\|_{\mathcal{H}_2}^2$. We also consider the effect of this choice considering also $K = 12$ and $K = 14$.

On one hand, we choose not to project the response variable before the test process, because some of the link between Y and X could be in the truncated part of Y . On the other hand, reducing the dimension for X is compulsory to solve the infinite dimension inverse problem. We consider the smoothing dimensions $q = 1$ and $q = 3$, with $h = n^{-1/(q+4)}$ for the test. Only the norm $\|\cdot\|_{\mathcal{H}_2}$ was used for the functional covariates. Our test rejects all the models when using $q = 1$. Meanwhile the model (V.8) is not rejected with $q = 3$. This could be explained by a possible lack of power due to smoothing in higher dimension.

References

- ANEIROS-PÉREZ, G. AND P. VIEU (2006): “Semi-functional partial linear regression,” *Statist. Probab. Lett.*, 76, 1102–1110.
- BOSQ, D. (2000): *Linear processes in function spaces*, vol. 149 of *Lecture Notes in Statist.*, Springer-Verlag, New York, theory and applications.
- CHIOU, J.-M. AND H.-G. MÜLLER (2007): “Diagnostics for functional regression via residual processes,” *Comput. Statist. Data Anal.*, 51, 4849–4863.
- DA PRATO, G. (2006): *An introduction to infinite-dimensional analysis*, Universitext, Springer-Verlag, Berlin, revised and extended from the 2001 original by Da Prato.
- DE JONG, P. (1987): “A central limit theorem for generalized quadratic forms,” *Probab. Theory Related Fields*, 75, 261–277.
- DELSOL, L., F. FERRATY, AND P. VIEU (2011): “Structural test in regression on functional variables,” *J. Multivariate Anal.*, 102, 422–447.
- FARAWAY, J. J. (1997): “Regression analysis for a functional response,” *Technometrics*, 39, 254–261.
- FERRATY, F., A. LAKSACI, A. TADJ, AND P. VIEU (2011): “Kernel regression with functional response,” *Electron. J. Stat.*, 5, 159–171.
- FERRATY, F., I. VAN KEILEGOM, AND P. VIEU (2012): “Regression when both response and predictor are functions,” *J. Multivariate Anal.*, 109, 10–28.
- GARCÍA-PORTUGUÉS, E., W. GONZÁLEZ-MANTEIGA, AND M. FEBRERO-BANDE (2012): “A goodness-of-fit test for the functional linear model with scalar response,” arXiv:1205.6167 [stat.ME].
- HÄRDLE, W. AND E. MAMMEN (1993): “Comparing nonparametric versus parametric regression fits,” *Ann. Statist.*, 21, 1926–1947.
- HORVÁTH, L. AND P. KOKOSZKA (2012): *Inference for functional data with applications*, Springer Ser. Statist., Springer, New York.
- KOKOSZKA, P., I. MASLOVA, J. SOJKA, AND L. ZHU (2008): “Testing for lack of dependence in the functional linear model,” *Canad. J. Statist.*, 36, 207–222.
- LAVERGNE, P., S. MAISTRE, AND V. PATILEA (2014): “A significance test for covariates in nonparametric regression.” arXiv:1403.7063 [math.ST].
- MAMMEN, E. (1993): “Bootstrap and wild bootstrap for high-dimensional linear models,” *Ann. Statist.*, 21, 255–285.
- PATILEA, V., C. SANCHEZ-SELLERO, AND M. SAUMARD (2012a): “Nonparametric testing for no-effect with functional responses and functional covariates,” arXiv:1209.2085 [math.ST].

- (2012b): “Projection-based nonparametric goodness-of-fit testing with functional covariates,” arXiv:1205.5578 [math.ST].
- RAMSAY, J. O. AND B. W. SILVERMAN (2005): *Functional data analysis*, Springer Ser. Statist., Springer, New York, second ed.
- VAN DER VAART, A. W. AND J. A. WELLNER (1996): *Weak convergence and empirical processes*, Springer Ser. Statist., New York: Springer-Verlag, with applications to statistics.
- (2011): “A local maximal inequality under uniform entropy,” *Electron. J. Stat.*, 5, 192–203.

VI Technical results and proofs

Proof of Lemma 2.1. The implication from left to right is obvious. For the reverse one, let us consider $l^2 \subset \mathbb{R}^\infty$ the space of real valued, square integrable sequences $x = (x_1, x_2, \dots)$, endowed with the scalar product $\langle x, y \rangle_2 = \sum_{k=1}^\infty x_k y_k$. Since any $w \in \mathcal{H}_2$ can be decomposed $w = \sum_{k \geq 1} \langle w, \phi_k \rangle_{\mathcal{H}_2} \phi_k$, where $\{\phi_1, \phi_2, \dots\}$ is the orthonormal basis considered in \mathcal{H}_2 , we shall use the usual identification between \mathcal{H}_2 and l^2 given by the isomorphism $w \in \mathcal{H}_2 \mapsto (\langle w, \phi_1 \rangle_{\mathcal{H}_2}, \langle w, \phi_2 \rangle_{\mathcal{H}_2}, \dots) \in l^2$. Denote $W_{12} = W_1 - W_2$.

Next, consider the linear operator Q from \mathcal{H}_2 into \mathcal{H}_2 defined by

$$Q\phi_k = a_k \phi_k, \quad k \geq 1.$$

The condition that the series $\sum_{k \geq 1} a_k$ is convergent means that the trace of the operator Q is finite. Now, since $\mathbb{E}[\|W_{12}\|_{\mathcal{H}_2}^2] < \infty$, there exists a set of events N such that $\mathbb{P}(N) = 1$ and for any $\omega \in N$, $W_{12}(\omega) \in l^2$ and hence $Q(W_{12}(\omega)) \in l^2$. By classical results in mathematical analysis in infinite-dimensional Hilbert spaces, see for instance Theorem 1.12 in Da Prato (2006), there exists a (unique) probability measure μ_Q on \mathcal{H}_2 endowed with the Borel σ -field such that for any $\omega \in N$,

$$\begin{aligned} \exp(-\|W_{12}(\omega)\|_{\mathcal{H}_2}^2/2) &= \exp(-\langle Q(W_{12}(\omega)), W_{12}(\omega) \rangle_{\mathcal{H}_2}/2) \\ &= \int_{\mathbb{R}^\infty} \exp(i \langle W_{12}(\omega), x \rangle_2) \mu_Q(x) \\ &= \int_{l^2} \exp(i \langle W_{12}(\omega), x \rangle_2) \mu_Q(x). \end{aligned}$$

The last equality expresses the fact that the probability measure μ_Q concentrates on l^2 . Using this identity for each $\omega \in N$, the inverse Fourier transform for $h^{-q}K((Z_1 - Z_2)/h)$, the Fubini theorem and a change of variables we can write

$$\begin{aligned} I(h) &= \int_{l^2} \int_{\mathbb{R}^q} \mathbb{E} \left[\left\langle V_1 e^{i\{t^\top Z_1 + \langle x, W_1 \rangle_2\}}, V_2 e^{-i\{t^\top Z_2 + \langle x, W_2 \rangle_2\}} \right\rangle_{\mathcal{H}_1} \right] \mathcal{F}[K](th) dt d\mu_Q(x) \\ &= \int_{l^2} \int_{\mathbb{R}^q} \left\| \mathbb{E} \left[V_1 e^{i\{t^\top Z_1 + \langle x, W_1 \rangle_2\}} \right] \right\|_{\mathcal{H}_1}^2 \mathcal{F}[K](th) dt d\mu_Q(x), \end{aligned}$$

where $V_j = \mathbb{E}[U_j \mid Z_j, W_j] \omega(Z_j)$, $j = 1, 2$. Deduce that

$$\mathbb{E} \left[\mathbb{E}[U \mid Z, W] \omega(Z) e^{i\{t^\top Z + \langle x, W \rangle_{\mathcal{H}_2}\}} \right] = 0, \quad \forall t \in \mathbb{R}^q, x \in l^2.$$

By the uniqueness of the Fourier Theorem in Hilbert spaces, see for instance Proposition 1.7 of Da Prato (2006), it follows that $\mathbb{E}[U \mid Z, W] = 0$. Now, the proof is complete. ■

Lemma VI.1 *Suppose that Assumptions D-(a) and K are met.*

(a)

$$\sup_{t \in \mathbb{R}^q} \frac{1}{nh^q} \sum_{i=1}^n K^k((t - Z_i)/h) = O_{\mathbb{P}} \left(\sqrt{\frac{\ln n}{nh^q}} \right) + o(h^{-q/2}),$$

for $k = 1$ or $k = 2$.

(b) Let $0 < \gamma_1, \gamma_2$ i.i.d. random variables such that $\mathbb{E}[\mathbb{E}^4(\gamma_1 \mid Z_1)] < \infty$. Then $\mathbb{E}[\gamma_1 \gamma_2 h^{-q} K^2((Z_1 - Z_2)/h)]$ converges to a positive constant as $h \rightarrow 0$.

Proof of Lemma VI.1. (a) We only consider the case $k = 1$, the case $k = 2$ is very similar. By Theorem 2.1 of van der Vaart and Wellner (2011),

$$\sup_{t \in \mathbb{R}^q} \left| \frac{1}{n} \sum_{i=1}^n \{K((t - Z_i)/h) - \mathbb{E}[K((t - Z)/h)]\} \right| = O_{\mathbb{P}} \left(\sqrt{\frac{h^q \ln n}{n}} \right) \quad (\text{VI.1})$$

Indeed, let \mathcal{G} be a class of functions of the observations with envelope function G , that here will be supposed bounded, and let

$$J(\delta, \mathcal{G}, L^2) = \sup_Q \int_0^\delta \sqrt{1 + \ln N(\varepsilon \|G\|_{Q,2}, \mathcal{G}, L^2(Q))} d\varepsilon, \quad 0 < \delta \leq 1,$$

denote the uniform entropy integral, where the supremum is taken over all finitely discrete probability distributions Q on the space of the observations, and $\|G\|_{Q,2}$ denotes the norm of G in $L^2(Q)$. Let Z_1, \dots, Z_n be a sample of independent observations and let

$$\mathbb{G}_n g = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{g(Z_i) - \mathbb{E}[g(Z)]\}, \quad \gamma \in \mathcal{G}$$

be the empirical process indexed by \mathcal{G} . If the covering number $N(\varepsilon, \mathcal{G}, L^2(Q))$ is of polynomial order in $1/\varepsilon$, there exists a constant $c > 0$ such that $J(\delta, \mathcal{G}, L^2) \leq c\delta\sqrt{\ln(1/\delta)}$ for $0 < \delta < 1/2$. Now if $\mathbb{E}g^2 < \delta^2 \mathbb{E}G^2$ for every γ and some $0 < \delta < 1$, Theorem 2.1 of van der Vaart and Wellner (2011) implies

$$\sup_{\mathcal{G}} |\mathbb{G}_n g| = J(\delta, \mathcal{G}, L^2) \left(1 + \frac{J(\delta, \mathcal{G}, L^2)}{\delta^2 \sqrt{n} \|G\|_2} \right) \|G\|_2 O_p(1), \quad (\text{VI.2})$$

where $\|G\|_2^2 = \mathbb{E}G^2$ and the $O_p(1)$ term is independent of n . Note that the family \mathcal{G} could change with n , as soon as the envelope is the same for all n . We can thus apply this result to the family of functions $\mathcal{G} = \{K((t - \cdot)/h) : t \in \mathbb{R}^q\}$ for a sequence h that converges to zero, the envelope $G(\cdot) \equiv K(0)$, and $\delta = h^{q/2}$. Its entropy number is of polynomial order in $1/\varepsilon$, independently of n , as $K(\cdot)$ is of bounded variation. Thus the rate in (VI.1) follows.

On the other hand, if $|\mathcal{F}[f_Z](u)|^{2-a}$ is integrable for some $a \in (0, 1]$, by the properties of the Fourier and inverse Fourier transforms, Fubini theorem and the Cauchy-Schwarz

inequality, for any $t \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{E}[h^{-q}K((t-Z)/h)] &= \left| (2\pi)^{-q/2} \mathbb{E} \int_{\mathbb{R}^n} \exp\{iu^\top t\} \exp\{-iu^\top Z\} \mathcal{F}[K](hu) du \right| \\
&= \left| \int_{\mathbb{R}^n} \exp\{iu^\top t\} \mathcal{F}[f_Z](u) \mathcal{F}[K](hu) du \right| \\
&\leq \left[\int_{\mathbb{R}^n} |\mathcal{F}[f_Z](u)|^{2-a} du \right]^{\frac{1}{2-a}} \left[\int_{\mathbb{R}} |\mathcal{F}[K](hu)|^{(2-a)/(1-a)} du \right]^{\frac{1-a}{2-a}} \\
&\leq C \left[h^{-q} \int_{\mathbb{R}^q} |\mathcal{F}[K](u)| du \right]^{\frac{1-a}{2-a}} \\
&= Ch^{-q(1-a)/(2-a)} \\
&= o(h^{-q/2}), \tag{VI.3}
\end{aligned}$$

for some constant C independent of t . Alternatively, if the density f_Z is bounded, by a change of variables we can write

$$\mathbb{E}[h^{-q}K((t-Z)/h)] = \int_{\mathbb{R}^q} K(u) f_Z(t-uh) du \leq C', \tag{VI.4}$$

for some constant C' independent of t . From equations (VI.12), (VI.1), (VI.3) and (VI.4)

$$\sup_{t \in \mathbb{R}^q} \frac{1}{nh^q} \sum_{i=1}^n K^k((t-Z_i)/h) = O_{\mathbb{P}} \left(\sqrt{\frac{\ln n}{nh^q}} \right) + o(h^{-q/2}), \tag{VI.5}$$

for $k = 1$ or $k = 2$.

(b) Let $e(z) = \mathbb{E}[\gamma_1 \mid Z_1 = z]$. If f_Z satisfies the condition $\int_{\mathbb{R}^q} |\mathcal{F}[f_Z]|^{2-a}(t) dt < \infty$ for some $a \in (0, 1]$, then $\int_{\mathbb{R}^q} f_Z^2 < \infty$ and hence by Cauchy-Schwarz inequality

$$\int_{\mathbb{R}^q} f_Z^2 e^2 \leq \left(\int_{\mathbb{R}^q} f_Z^2 \right)^{1/2} \left(\int_{\mathbb{R}^q} f_Z e^4 \right)^{1/2} < \infty.$$

Using Fubini Theorem, the inverse Fourier Transform formula and Parseval identity, we can write

$$\begin{aligned}
\mathbb{E}[\gamma_1 \gamma_2 h^{-q} K((Z_1 - Z_2)/h)] &= \mathbb{E}[h^{-q} e(Z_1) e(Z_2) K((Z_1 - Z_2)/h)] \\
&= (2\pi)^{-\frac{q}{2}} \mathbb{E} \int_{\mathbb{R}^n} e(Z_1) \exp\{iu^\top Z_1\} \\
&\quad \times e(Z_2) \exp\{-iu^\top Z_2\} \mathcal{F}[K](hu) du \\
&= \int_{\mathbb{R}^n} |\mathcal{F}[f_Z e](u)|^2 |\mathcal{F}[K](hu)| du \\
&\rightarrow \int_{\mathbb{R}^n} |\mathcal{F}[f_Z e](u)|^2 du \\
&= \int_{\mathbb{R}^n} f_Z^2(u) e^2(u) du,
\end{aligned}$$

where for the limit we use the Dominated Convergence Theorem. If f_Z is bounded, we can use a change of variables like for equation (VI.4) and again the Dominated Convergence Theorem to obtain the same strictly positive and finite limit. ■

Proof of Theorem 3.1. The proof is based on the Central Limit Theorem 5.1 of de Jong (1987). Let

$$\Omega_{ij} = \frac{1}{n(n-1)h^q} \langle U_i \omega(Z_i), U_j \omega(Z_j) \rangle_{\mathcal{H}_1} K_{ij}(h) \varphi_{ij}, \quad 1 \leq i \neq j \leq n,$$

and $\Omega_{ii} = 0$, $1 \leq i \leq n$. Let $\Omega(n) = \sum_{i \neq j} \Omega_{ij}$ and $\sigma(n)^2 = 2 \sum_{i \neq j} \sigma_{ij}^2$ where

$$\sigma_{ij}^2 = \mathbb{E}[\Omega_{ij}^2 \mid (Z_i, W_i), (Z_j, W_j)] = \frac{V_{ij}^2 K_{ij}^2(h) \varphi_{ij}^2}{n^2(n-1)^2 h^{2q}}$$

and

$$V_{ij}^2 = \mathbb{E}[\langle U_i \omega(Z_i), U_j \omega(Z_j) \rangle_{\mathcal{H}_1}^2 \mid (Z_i, W_i), (Z_j, W_j)]$$

Consider the following conditions:

1. there exists a sequence of real numbers k_n such that

$$k_n^2 \sigma(n)^{-2} \max_{1 \leq i \leq n} \sum_{1 \leq j \leq n} \sigma_{ij}^2 = o_{\mathbb{P}}(1) \quad (\text{VI.6})$$

and

$$\max_{1 \leq i \neq j \leq n} \sigma_{ij}^{-2} \mathbb{E}[\Omega_{ij}^2 \mathbf{1}_{\{|\Omega_{ij}| > k_n \sigma_{ij}\}} \mid (Z_i, W_i), (Z_j, W_j)] = o_{\mathbb{P}}(1); \quad (\text{VI.7})$$

- 2.

$$\sigma(n)^{-2} \max_{1 \leq i \leq n} \mu_i^2 = o_{\mathbb{P}}(1), \quad (\text{VI.8})$$

where μ_1, \dots, μ_n are the eigenvalues of the matrix (σ_{ij}) .

If these conditions hold true, using the characterization of the convergence in probability based on almost surely convergence subsequences, Theorem 5.1 of de Jong (1987) applied conditionally on the covariates implies that for any $t \in \mathbb{R}$,

$$\mathbb{P}(\sigma(n)^{-1} \Omega(n) \leq t \mid (Z_1, W_1), \dots, (Z_n, W_n)) - \Phi(t) = o_{\mathbb{P}}(1).$$

By the dominated convergence theorem, $\sigma(n)^{-1} \Omega(n)$ converges in law to a standard normal distribution. Hence, it remains to check conditions (VI.6) to (VI.8).

First, let us bound from below $\sigma(n)$. By Assumption D-(c)-(i), $V_{ij}^2 \geq \underline{\sigma}^2$ almost surely, so that

$$\sigma(n)^2 \geq \frac{\underline{\sigma}^2}{n^2(n-1)^2 h^{2q}} \sum_{i \neq j} K_{ij}^2(h) \varphi_{ij}^2 \geq \frac{\underline{\sigma}^2}{n^2(n-1)^2 h^{2q}} \sum_{i \neq j} K_{ij}^2(h) \lambda_i^2 \lambda_j^2 = \underline{\sigma}^2 \tau(n)^2,$$

where $\lambda_i = \exp(-2\|W_i\|_{\mathcal{A}}^2)$. By standard calculations, the variance $n(n-1)h^q\tau(n)^2$ tends to zero. By Lemma VI.1-(b) the expectation of $n(n-1)h^q\tau(n)^2$ tends to a positive constant. Deduce that

$$\sigma(n)^{-2} = O_{\mathbb{P}}(n^2h^q). \quad (\text{VI.9})$$

Next, note that by Hölder inequality and Assumption D-(c)-(ii),

$$V_{ij}^2 \leq \mathbb{E}[\|U_i\omega(Z_i)\|_{\mathcal{H}_1}^2 \mid (Z_i, W_i), (Z_j, W_j)] \mathbb{E}[\|U_j\omega(Z_j)\|_{\mathcal{H}_1}^2 \mid (Z_i, W_i), (Z_j, W_j)] \leq C^{4/\nu},$$

almost surely. Deduce from this and Lemma VI.1-(a) that

$$\max_{1 \leq i \leq n} \sum_{1 \leq j \leq n} \sigma_{ij}^2 = o_{\mathbb{P}}(n^{-2}h^{-q}). \quad (\text{VI.10})$$

Then condition (VI.6) follows from (VI.9) and (VI.10) for some suitable sequence $k_n \rightarrow \infty$.

Next, let us note that

$$\frac{\Omega_{ij}^2}{\sigma_{ij}^2} \leq \underline{\sigma}^{-2} \langle U_i\omega(Z_i), U_j\omega(Z_j) \rangle_{\mathcal{H}_1}^2,$$

so that for any i and j , by Hölder inequality, Markov inequality and Assumption D-(c)-(ii),

$$\begin{aligned} \sigma_{ij}^{-2} \mathbb{E} [\Omega_{ij}^2 \mathbf{1}_{\{\Omega_{ij} > k_n \sigma_{ij}\}} \mid (Z_i, W_i), (Z_j, W_j)] &\leq \underline{\sigma}^{-2} \mathbb{E} [\langle U_i\omega(Z_i), U_j\omega(Z_j) \rangle_{\mathcal{H}_1}^2 \\ &\quad \times \mathbf{1}_{\{\langle U_i\omega(Z_i), U_j\omega(Z_j) \rangle_{\mathcal{H}_1} > \underline{\sigma} k_n\}} \mid (Z_i, W_i), (Z_j, W_j)] \\ &\leq \underline{\sigma}^{-2} \mathbb{E}^{2/\nu} [\|U_i\omega(Z_i)\|_{\mathcal{H}_1}^\nu \|U_j\omega(Z_j)\|_{\mathcal{H}_1}^\nu \mid (Z_i, W_i), (Z_j, W_j)] \\ &\quad \times \mathbb{P}^{(\nu-2)/\nu} [\langle U_i\omega(Z_i), U_j\omega(Z_j) \rangle_{\mathcal{H}_1} > \underline{\sigma} k_n \mid (Z_i, W_i), (Z_j, W_j)] \\ &\leq \underline{\sigma}^{-2} C^{4/\nu} [C^{2/\nu} (\underline{\sigma} k_n)^{-1}]^{(\nu-2)/\nu}, \end{aligned}$$

almost surely. Thus condition (VI.7) holds true for any $k_n \rightarrow \infty$.

To check condition (VI.8), let \mathcal{K} be the matrix with elements

$$\mathcal{K}_{ij} = V_{ij} \varphi_{ij} K((Z_i - Z_j)/h) / [n(n-1)h^q], \quad i \neq j, \quad \text{and} \quad \mathcal{K}_{ii} = 0, \quad (\text{VI.11})$$

and $\|\mathcal{K}\|_2$ is the spectral norm of \mathcal{K} . By definition, $\|\mathcal{K}\|_2 = \sup_{u \in \mathbb{R}^n, u \neq 0} \|\mathcal{K}u\| / \|u\|$ and $|u' \mathcal{K} w| \leq \|\mathcal{K}\|_2 \|u\| \|w\|$ for any $u, w \in \mathbb{R}^n$. By Cauchy-Schwarz inequality, for any $u \in \mathbb{R}^n$,

$$\begin{aligned} \|\mathcal{K}u\|^2 &= \sum_{i=1}^n \left(\sum_{j=1, j \neq i}^n V_{ij} \frac{K((Z_i - Z_j)/nh)}{h^q n(n-1)} u_j \right)^2 \\ &\leq C^{4/\nu} \sum_{i=1}^n \left(\sum_{j=1, j \neq i}^n \frac{K((Z_i - Z_j)/nh)}{h^q n(n-1)} \right) \sum_{j=1, j \neq i}^n \frac{K((Z_i - Z_j)/nh)}{h^q n(n-1)} u_j^2 \\ &\leq C^{4/\nu} \|u\|^2 n^{-2} \left[\max_{1 \leq i \leq n} \left(\sum_{j=1, j \neq i}^n \frac{K((Z_i - Z_j)/nh)}{h^q (n-1)} \right) \right]^2, \end{aligned} \quad (\text{VI.12})$$

for some constant $c > 0$. By Lemma VI.1-(a) deduce that

$$\max_{1 \leq i \leq n} \mu_i^2 = \frac{1}{n^2} \left[O_{\mathbb{P}} \left(\frac{\ln n}{nh^q} \right) + o(h^{-q}) \right] = o_{\mathbb{P}} \left(\frac{1}{n^2 h^q} \right).$$

Condition (VI.8) follows from this and the rate (VI.9). Now the proof is complete. ■

Proof of Theorem 3.2. Let us simplify the notation and denote $\omega_i = \omega(Z_i)$ and $\delta_i = \delta(Z_i, W_i)$. Next let us decompose

$$\begin{aligned} I_n(h) &= \frac{1}{n(n-1)h^q} \sum_{1 \leq i \neq j \leq n} \langle U_i^0 \omega_i, U_j^0 \omega_j \rangle_{\mathcal{H}_1} K_{ij}(h) \varphi_{ij} \\ &+ \frac{r_n}{n(n-1)h^q} \sum_{1 \leq i \neq j \leq n} \langle U_i^0 \omega_i, \delta_j \omega_j \rangle_{\mathcal{H}_1} K_{ij}(h) \varphi_{ij} \\ &+ \frac{r_n^2}{n(n-1)h^q} \sum_{1 \leq i \neq j \leq n} \langle \delta_i \omega_i, \delta_j \omega_j \rangle_{\mathcal{H}_1} K_{ij}(h) \varphi_{ij} \\ &= I_{0n} + 2I_{1n} + I_{2n}. \end{aligned}$$

The rate of I_{0n} is given by Theorem 3.1, so that it remains to investigate the rates of I_{1n} and I_{2n} and to bound in probability $v_n^2(h)$. By standard calculations,

$$\mathbb{E}(I_{0n}) = \mathbb{E}(I_{1n}) = 0, \quad \text{Var}(I_{0n}) = O_{\mathbb{P}}(n^{-2}h^{-q}) \quad \text{and} \quad \text{Var}(I_{1n}) = O_{\mathbb{P}}(r_n^2 n^{-1}h^{-q}).$$

Moreover,

$$\mathbb{E}(I_{2n}) = r_n^2 \mathbb{E}[\langle \delta_1 \omega_1, \delta_2 \omega_2 \rangle_{\mathcal{H}_1} h^{-q} K_{12}(h) \varphi_{12}] \quad \text{and} \quad \text{Var}(I_{1n}) = O_{\mathbb{P}}(r_n^4 n^{-1}h^{-q}).$$

By dominated convergence we have

$$\begin{aligned} \mathbb{E}[\langle \delta_1 \omega_1, \delta_2 \omega_2 \rangle_{\mathcal{H}_1} h^{-q} K_{12}(h) \varphi_{12}] &= \mathbb{E}[\mathbb{E}[\langle \delta_1 \omega_1, \delta_2 \omega_2 \rangle_{\mathcal{H}_1} \varphi_{12} \mid Z_1, Z_2] h^{-q} K_{12}(h)] \\ &= (2\pi)^{-\frac{q}{2}} \mathbb{E} \left[\int_{\mathbb{R}^n} \mathbb{E}[\langle \delta_1 \omega_1, \delta_2 \omega_2 \rangle_{\mathcal{H}_1} \varphi_{12} \mid Z_1, Z_2] \right. \\ &\quad \times \exp\{iu^\top Z_1\} \exp\{-iu^\top Z_2\} \mathcal{F}[K](hu) du \Big] \\ &\rightarrow (2\pi)^{-\frac{q}{2}} \int_{\mathbb{R}^n} \mathbb{E}[\langle \delta_1 \omega_1, \delta_2 \omega_2 \rangle_{\mathcal{H}_1} \varphi_{12} \\ &\quad \times \exp\{iu^\top Z_1\} \exp\{-iu^\top Z_2\}] du \end{aligned}$$

By arguments as used in the proof of Lemma 2.1 the expectation of I_{2n} could be shown to be strictly positive. Since $r_n^2 n h^{q/2} \rightarrow \infty$, the result follows. ■

Lemma 6.1 *Suppose the assumptions of Theorem 4.1 hold true. Then, for $l = 1$ and $l = 2$,*

$$\max_{1 \leq i \leq n} \frac{1}{nh} \sum_{j=1}^n \left| [\widehat{K}_{ij}(h) \widehat{\varphi}_{ij}]^l - [K_{ij}(h) \varphi_{ij}]^l \right| = o_{\mathbb{P}}(1) \quad (\text{VI.13})$$

and

$$\frac{1}{n(n-1)h} \sum_{1 \leq i \neq j \leq n} \left\{ [\widehat{K}_{ij}(h)\widehat{\varphi}_{ij}]^2 - [K_{ij}(h)\varphi_{ij}]^2 \right\} = o_{\mathbb{P}}(1). \quad (\text{VI.14})$$

Proof of Lemma 6.1. Assume that $nh^4/\ln^2 n \rightarrow \infty$. By the Lipschitz property of the kernel and of the $\varphi(\cdot)$ function, the bound (IV.2) and conditions (IV.3), for $l = 1$ and $l = 2$,

$$\max_{i,j} h^{-1} \left| [\widehat{K}_{ij}(h)\widehat{\varphi}_{ij}]^l - [K_{ij}(h)\varphi_{ij}]^l \right| \leq Ch^{-2}\Delta_n \max_{1 \leq i \leq n} |\Gamma_i| = O_{\mathbb{P}}(n^{-1/2}h^{-2} \ln n) = o_{\mathbb{P}}(1).$$

The rates (VI.13) and (VI.14) follow.

If conditions at point (2) or point (3) are met, the arguments are of different nature. First, note that the conditions of point (3) involve that f_Z is bounded. Next, since the kernel K is of bounded univariate kernels, let K_1 and K_2 non decreasing bounded functions such that $K^l = K_1 - K_2$ and denote $K_{1h} = K_1(\cdot/h)$. Clearly, it is sufficient to prove (VI.13) for K_1 , similar arguments apply for K_2 and hence we get the results for K^l . For simpler writings let us assume that K^l is differentiable and let $K_1(x) = \int_{-\infty}^x |K^{l'}(t)|dt$, $x \in \mathbb{R}$. The general case of a bounded variation K^l can be handled with obvious modifications. Let $b_n \downarrow 0$ such that $b_n\sqrt{n}/\ln n \rightarrow \infty$ and define the event

$$\mathcal{E}_{1n} = \left\{ \max_{1 \leq i \leq n} [\|\widehat{Z}_i - Z_i\| + \|\widehat{W}_i - W_i\|_{\mathcal{H}_2}] \leq b_n \right\}, \quad (\text{VI.15})$$

so that $\mathbb{P}(\mathcal{E}_{1n}^c) \rightarrow 0$. Since $|\exp(-t^2/2) - \exp(-s^2/2)| \leq |t - s|$, on the set \mathcal{E}_{1n} , $\forall i, j$

$$\begin{aligned} & -b_n K_{1h}(Z_i - Z_j + 2b_n) - [K_{1h}(Z_i - Z_j - 2b_n) - K_{1h}(Z_i - Z_j)]\varphi_{ij} \\ & \leq \left| K_{1h}(\widehat{Z}_i - \widehat{Z}_j) - K_{1h}(Z_i - Z_j) \right| \varphi_{ij} - K_{1h}(\widehat{Z}_i - \widehat{Z}_j) |\varphi_{ij} - \widehat{\varphi}_{ij}| \\ & \leq \left| K_{1h}(\widehat{Z}_i - \widehat{Z}_j) \widehat{\varphi}_{ij} - K_{1h}(Z_i - Z_j) \varphi_{ij} \right|. \end{aligned}$$

Similarly

$$\begin{aligned} & \left| K_{1h}(\widehat{Z}_i - \widehat{Z}_j) \widehat{\varphi}_{ij} - K_{1h}(Z_i - Z_j) \varphi_{ij} \right| \\ & \leq b_n K_{1h}(Z_i - Z_j + 2b_n) + [K_{1h}(Z_i - Z_j + 2b_n) - K_{1h}(Z_i - Z_j)]\varphi_{ij}. \end{aligned}$$

We focus on the second inequality, the first one can be handled similarly. To justify

(VI.13) we can write

$$\begin{aligned}
& \frac{1}{nh} \sum_{j=1}^n \left| K_{1h} \left(\widehat{Z}_i - \widehat{Z}_j \right) \widehat{\varphi}_{ij} - K_{1h} (Z_i - Z_j) \varphi_{ij} \right| \\
& \leq \frac{b_n}{h} \mathbb{E}[K_{1h} (Z_i - Z + 2b_n) \mid Z_i] \\
& \quad + \frac{b_n}{nh} \sum_{j=1}^n [K_{1h} (Z_i - Z_j + 2b_n) - \mathbb{E}[K_{1h} (Z_i - Z_j + 2b_n) \mid Z_i]] \\
& \quad + \frac{1}{nh} \sum_{j=1}^n \{K_{1h} (Z_i - Z_j + 2b_n) \varphi_{ij} - \mathbb{E}[K_{1h} (Z_i - Z_j + 2b_n) \varphi_{ij} \mid Z_i, W_i]\} \\
& \quad - \frac{1}{nh} \sum_{j=1}^n \{K_{1h} (Z_i - Z_j) \varphi_{ij} - \mathbb{E}[K_{1h} (Z_i - Z_j) \varphi_{ij} \mid Z_i, W_i]\} \\
& \quad + \mathbb{E}[h^{-1} K_{1h} (Z_i - Z_j + 2b_n) \varphi_{ij} \mid Z_i, W_i] - \mathbb{E}[h^{-1} K_{1h} (Z_i - Z_j) \varphi_{ij} \mid Z_i, W_i] \\
& = A_{0n,i} + A_{1n,i} + A_{2n,i} - A_{3n,i} + A_{4n,i}.
\end{aligned}$$

Since f_Z is supposed bounded, by a simple change of variable, we get $\max_{1 \leq i \leq n} A_{0n,i} = O(b_n) = o(1)$.

The terms $A_{1n,i}$, $A_{2n,i}$ and $A_{3n,i}$ could be treated similarly, hence we only investigate $A_{2n,i}$. First note that, since $\varphi_{ij} \leq 1$, the function K_1 is bounded and integrable and f_Z is bounded, for all j we have

$$\text{Var} (K_{1h} (Z_i - Z_j + 2b_n) \varphi_{ij} \mid Z_i, W_i) \leq Ch, \quad \forall i,$$

for some constant C independent of n and Z_i, W_i . Using this conditional variance bound and applying Bernstein inequality* conditionally on the Z_i, W_i 's, for any $t > 0$

$$\begin{aligned}
& \mathbb{P} \left[\max_{1 \leq i \leq n} \left| \frac{1}{n} \sum_{j=1}^n \{K_{1h} (Z_i - Z_j + 2b_n) - \mathbb{E}[K_{1h} (Z_i - Z_j + 2b_n) \mid X_i, W_i]\} \right| > th \right] \\
& \leq \sum_{i=1}^n \mathbb{E} \left[\mathbb{P} \left[\left| \frac{1}{n} \sum_{j=1}^n K_{1h} (Z_i - Z_j + 2b_n) - \mathbb{E}[K_{1h} (Z_i - Z_j + 2b_n) \mid X_i, W_i] \right| > th \mid X_i, W_i \right] \right] \\
& \leq 2n \exp \left(-\frac{t^2}{2} \frac{nh^2}{Ch + thM/3} \right) \leq 2 \exp \left([\ln n] \left[1 - \frac{t^2}{2} \frac{nh/\ln n}{C + tM/3} \right] \right) \rightarrow 0,
\end{aligned}$$

since $nh/\ln n \rightarrow \infty$ under the conditions of point (2) or those of point (3) (here M is any constant that bounds K_1). Deduce that $\max_{1 \leq i \leq n} A_{2n,i} = o_{\mathbb{P}}(1)$.

*Recall that Bernstein inequality states that if W_1, \dots, W_n are i.i.d. centered random variables of variance σ^2 taking values in the interval $[-M, M]$, then for any $s > 0$

$$\mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n W_i \right| > s \right) \leq 2 \exp \left(-\frac{ns^2}{2[\sigma^2 + Ms/3]} \right).$$

To complete the proof of (VI.13) it remains to investigate the convergence of $A_{4n,i}$ uniformly with respect to i . First, since $\varphi_{ij} \leq 1$,

$$\begin{aligned} & |\mathbb{E}[h^{-1}K_{1h}(Z_i - Z_j + 2b_n)\varphi_{ij} \mid Z_i, W_i] - \mathbb{E}[h^{-1}K_{1h}(Z_i - Z_j)\varphi_{ij} \mid Z_i, W_i]| \\ & \leq \mathbb{E}[|h^{-1}K_{1h}(Z_i - Z_j + 2b_n) - h^{-1}K_{1h}(Z_i - Z_j)| \mid Z_i]. \end{aligned}$$

If the conditions of point (2) are met, suppose that the sequence b_n used for the definition of the set \mathcal{E}_{1n} in equation (VI.15) is such that $b_n/h \rightarrow 0$ and $b_n\sqrt{n}/\ln n \rightarrow \infty$. By obvious calculations and changes of variables and using the uniform bound for f_Z , for each $1 \leq i \leq n$

$$\begin{aligned} & \mathbb{E}[|h^{-1}K_{1h}(Z_i - Z_j + 2b_n) - h^{-1}K_{1h}(Z_i - Z_j)| \mid Z_i] \\ & = \int_{\mathbb{R}} [K_1(u + 2b_nh^{-1}) - K_1(u)] f_Z(Z_i - uh) du \\ & \leq \int_{\mathbb{R}} \left[\int_0^{2b_nh^{-1}} |K'(u+v)| dv \right] f_Z(Z_i - uh) du \\ & \leq C \int_0^{2b_nh^{-1}} \left[\int_{\mathbb{R}} |K'(u+v)| du \right] dv \rightarrow 0. \quad (\text{VI.16}) \end{aligned}$$

Finally, if the conditions of point (3) are met, by an alternative but still obvious change of variables and using the uniform continuity of f_Z and the fact that $b_n \rightarrow 0$, for each $1 \leq i \leq n$

$$\begin{aligned} & \mathbb{E}[|h^{-1}K_{1h}(Z_i - Z_j + 2b_n) - h^{-1}K_{1h}(Z_i - Z_j)| \mid Z_i] \\ & \leq \int_{\mathbb{R}} K_1(u) |f_Z(2b_n + Z_i - uh) - f_Z(Z_i - uh)| du \\ & \leq \sup_{t \in \mathbb{R}} |f_Z(2b_n + t) - f_Z(t)| \int_{\mathbb{R}} K_1(u) du \rightarrow 0. \end{aligned}$$

Now the arguments are complete for justifying (VI.13). The rate in (VI.14) could be easily derived from the rate in (VI.13). ■

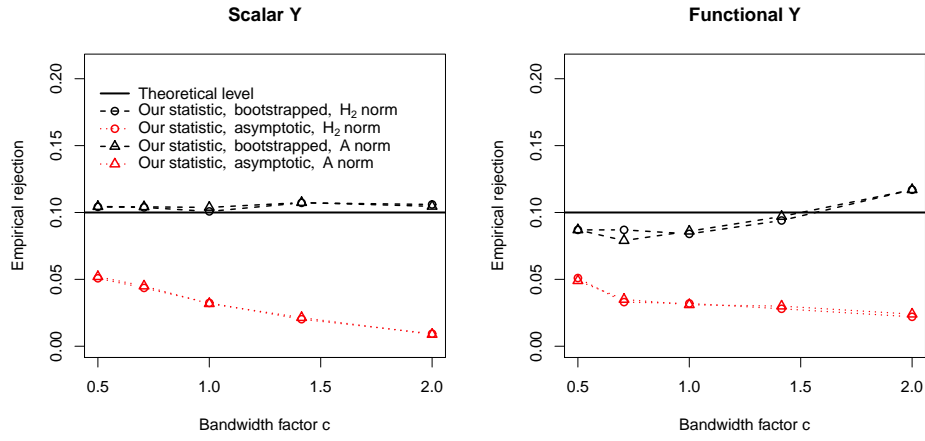


Figure 1: Empirical rejection for scalar (left panel) and functional (right panel) Y under the null hypothesis.

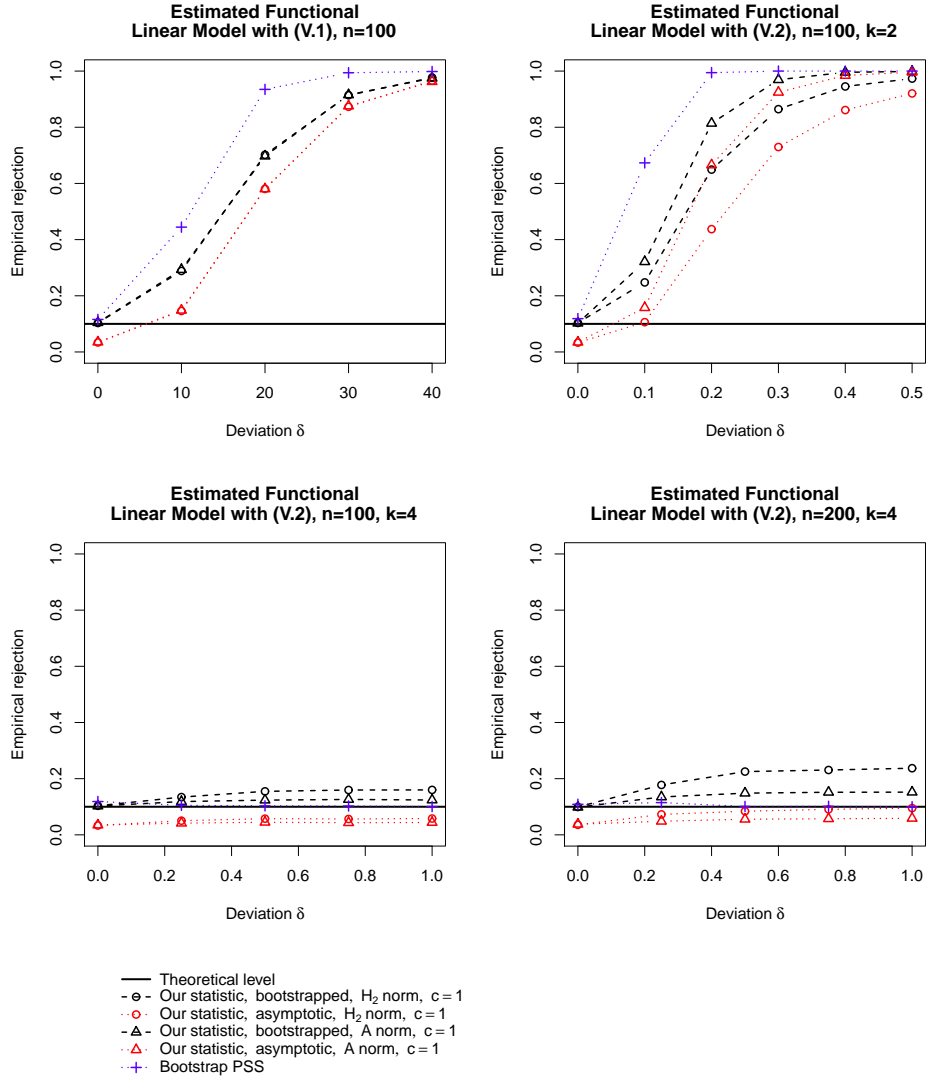


Figure 2: Empirical rejection for scalar Y .

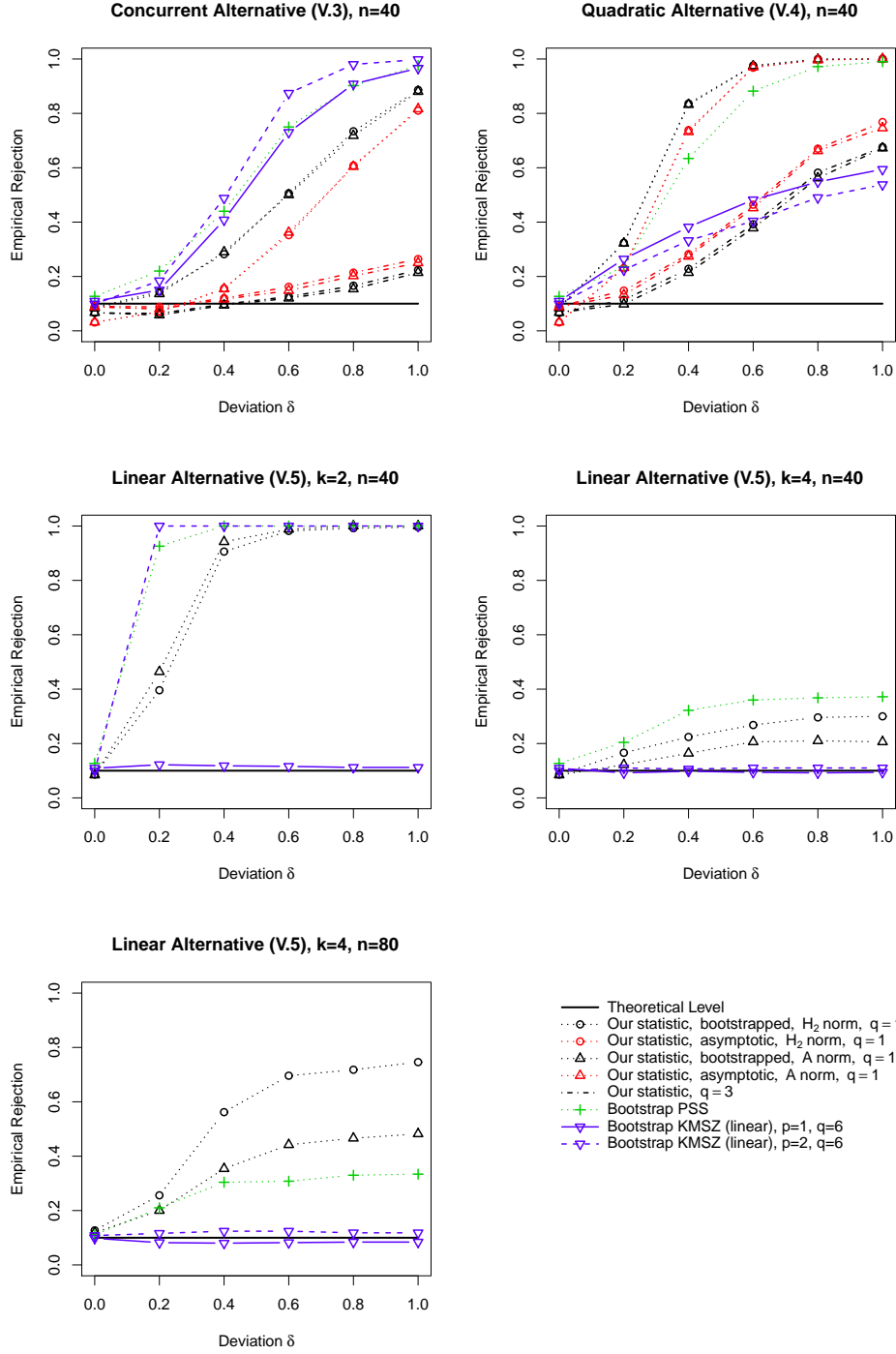


Figure 3: Empirical rejection for functional Y .